SCIENCE

FRIDAY, DECEMBER 20, 1895.

CONTENTS:

MSS. intended for publication and books, etc., intended for review should be sent to the responsible editor, Prof. J. McKeen Cattell, Garrison on Hudson, N. Y.
Subscriptions and advertisements should be sent to SCIENCE, 41 N. Queen St., Lancaster, Pa., or 41 East 49th St., New York.

THE LAW OF THE LONG RUN.

"MEN were surprised to hear that not only births, deaths and marriages, but the decisions of tribunals, the results of popular elections, the influence of punishments in checking crime, the comparative values of medical remedies, the probable limits of error in numerical results in every department of physical inquiry, the detection of causes, physical, social and moral, nay even the weight of evidence and the validity of logical argument, might come to be surveyed with the lynx-eyed scrutiny of a dispassionate analysis."

So wrote Sir John Herschel, a good many years ago, of the *Calculus of Probabilities*, which had just come into prominence through important practical applications. The 'Doctrine of Chance' is apparently miscalled because it is chiefly applied to the study and development of natural laws in the operation of which there can be no such thing as chance.

Popularly the word 'chance' is often used as if to imply the *absence* of any cause, but this is an unreasonable, if not an unthinkable condition. Really such words as 'chance,' 'accident' and the like imply only the absence of any assigned or recognized cause, and the doctrine of chances is a study and development of the laws relating to a series or aggregation of events, concerning the individual components of which we are absolutely ignorant. Thus, if

one tosses a coin, it is, in general, impossible to know in advance on which face it will rest. That its behavior in this respect will be governed by the operation of forces and conditions, just as certain and just as definitely compelling a given result as is the behavior of the sun and moon in the matter of an eclipse, will not be denied. If in any particular trial we knew all of the forces and conditions which influenced the result we should find that they were never equally balanced between the two possible events, but always predominated in favor of that which actually happened. A complete knowledge of antecedent causes would reveal the fact that each of these (to us at present unknown) forces and conditions is subject to other secondary influences which continually change its resultant effect from one side to the other, and so on, in lower degree, to the end *that in a very large number of trials* the ratio of the number of times the two possible events have occurred becomes very nearly one, to which, indeed, it approximates continuously as the number of trials increases. Note the use of the word *ratio* in this statement. In a very large number of trials in tossing a coin the number of heads may be always in excess of the tails and by a continually increasing amount, and yet the ratio of the two may be continually tending towards equality. It is important to call attention to the dependence of the Theory of Chances upon experience and experiment. It is not rigorously true, as is often stated in writing about probabilities, that if a coin is tossed in the air "it is as likely to fall upon one face as another." Such a condition necessitates an absolutely equal division of all forces and conditions between the two possible events, and it is logical to conclude that neither would happen. A more nearly correct statement would be that we are quite ignorant of any cause tending to one result rather than to another. As there are, apparently, but two possible results we may put the *a priori* probability of each at one-half. This conclusion is, however, of little value until experience has proved that in the case under consideration the controlling forces and conditions are so evenly distributed and nicely adjusted that the balance is easily thrown from one side to the other. If experience shows that in a certain series of trials there is a marked tendency for a coin to fall on one face rather than the other we are led at once to suspect that there is something in the manner of tossing, or in the nature of the coin itself, or in some other less easily understood condition, which has caused this tendency, and we know that our numerical expression for the probability cannot be correct. Experience, therefore, is essential to any useful application of this doctrine, and experience is valuable only when it is large. The numerical evaluation of a probability must be, at least, one which is not contradictory to experience and, whenever possible, it must be one supported and verified by experiment.

The above general remarks on the Doctrine of Chances (with apologies to the many who are quite familiar with the subject) are submitted with a desire to aid in clearing away some of the difficulties which many people encounter in trying to understand the usefulness of this most interesting branch of applied mathematics. In scientific investigation whenever our knowledge is so nearly complete and our mental vision so far reaching that we can trace the progress of the phenomenon under consideration, or of each of its elements from beginning to end, we do not need its aid. In the thousands of instances, however, in which primary causes are so obscure and so numerous that we can only know them by their integrated effects, its assistance has proved to be of incalculable value.

The object of the present article is to remind the reader that whenever the number

of these elementary controlling forces and conditions (generally quite unknown as such) is sufficiently large there will be a *definite integral*, which becomes more stable in form and character as the number from which it is derived increases, and that it may be depended upon and treated with as much confidence as if it were an observed and explained phenomenon. This is, of course, the basis of all statistical studies of natural phenomena. One or two simple illustrations may be given. In the case of tossing a coin it may be impossible to discover by any physical examination of the coin itself, or of the conditions influencing it when thrown in the air, any reason for the appearance of one face rather than the other. The *a priori* probability of the appearance of a given face may, therefore, be properly put at one-half. But in ten thousand trials there might be shown a tendency towards the appearance of heads, and if this persisted with an increase of the number of trials it would be legitimate to conclude that the coin was not uniform or symmetrical in structure or that the balance of forces and condition in tossing was not good. The universality of this principle has given rise to the idea of the *long run*, or, as it is sometimes put, the Law of the Long Run. In simple language this means that however obscure or relatively ineffective an influencing condition may be, *in the long run* it will make itself felt and may be evaluated in quantity and character if the number of examples is sufficiently great. It will be observed that this principle is different from, although not necessarily inconsistent with, the statement often made that minor departures from a general law, due to minute and continually varying influences, will, 'in the long run,' cancel and destroy each other.

It is not necessary to quote examples of the useful aplication of this principle of the long run in bringing to light hitherto unsuspected relations or unconsidered influences, but I may be allowed to refer to one simple and easily understood illustration, an account of which was published about ten years ago. It was founded on the following reasoning: An author with a generous vocabulary at his service must be continually making a choice among words that are nearly identical in meaning. The influences which control the choice are often numerous and doubtless generally unrecognized, but in the long run a certain set will prevail and the composition will be marked by this characteristic. It might not be impossible to discover the existence of each separate influence by an extensive analysis of the author's composition in such a manner as to reveal the characteristic of this influence to the exclusion of all others, but the labor of doing this would in many cases be enormous. In the paper referred to, the simple and easily reached characteristic dependent on the number of letters in the words used was proposed, and it was shown that when properly analyzed the composition of any author could be made to produce what was called a 'characteristic curve,' which, it was suggested, might prove to be peculiar to him and which might thus afford a clue to his identification. Some further applications of the suggested method have been made since the time of first publication which have tended strongly to confirm the view then held.

The application of the calculus of probabilities to the determination of life expectation and other quantities of great importance to life insurance and annuity companies has long been admitted, and statistical methods based on the principle of the long run have long been in vogue in the study of the distribution and prevalence of disease. There is good reason for believing that what is ordinarily known as purely accidental death and injury is governed in distribution by the same inexorable laws.

For illustrations of this proposition I am indebted to an officer of one of the great railway systems of the country, who has kindly furnished, during the past three or four years, most interesting statistical information relating to accidents, collected by him with the object of studying the results in the interests of the corporation with which he is connected. I have put the principal results in graphic form, but for those who like to see the numbers from which the diagrams were constructed I have included tables showing the classification of accidents, as to occupation, results, etc. Four years are included in this investigation, and there is shown in the tables and diagrams the average for the whole period.

on the track or other property of the corporation. The meaning of the diagrams will be readily seen, the circle, in each case, being divided into sectors proportional to the number of accidents in the several classes, the whole area representing the total in every instance, *regardless of the numerical magnitude of that total*, approximate constancy of *ratio* of distribution being the point under consideration. The persistency of this ratio is certainly very striking. Naturally the railway corporation collects this information with the view of being benefited by it, and therefore it may be expected that, as its character is developed from year to year, the operation of what has been called 'chance' in controlling the distribution of accidents will be

TABLE A.—CLASSIFIED AS TO OCCUPATION.

	NUMBER OF CASUALTIES.					RATIO OF DISTRIBUTION.				
	1891.	1892.	1893.	1894.	Mean.	1891.	1892.	1893.	1894.	Mean.
						%	%	%	%	%
Passengers	257	221	281	130	222	7.6	5.4	6.6	4.2	6.0
Travelers on public highway	190	239	179	195	201	5.6	5.9	4.6	6.3	5.6
Employés	2488	3105	3087	2339	2755	73.4	76.6	77.4	75.7	75.7
Trespassers	455	492	447	430	456	13.4	12.1	11.4	13.8	12.7
Total	3390	4057	3994	3094	3634					

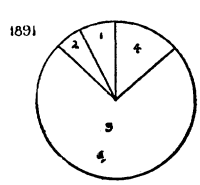

FIG. 1.

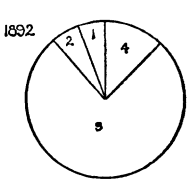

FIG. 2.

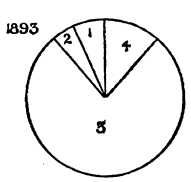

FIG. 3.

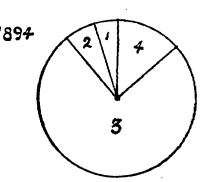

FIG. 4.

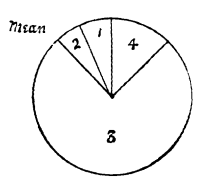

FIG. 5.

OCCUPATION.

1. Passengers. 2. Travelers on highway. 3. Employés. 4. Trespassers.

The first five diagrams show the distribution of accidents among the various occupations of the injured, at the time of the injury. The division is into the four general classes of passengers, employés, travelers on the public highways and trespassers

interfered with by new influences which will tend strongly to diminish the number of casualties, especially in those classes in which accident is most costly to the corporation. The results exhibited herewith furnish evidence that this influence is already felt.

It might be claimed that this constancy of ratio of distribution of casualties among the four classes is only a reflection of the constancy of the ratio of the numbers of those classes. It will be noted, however, that in the case of only two of them can anything be known of that ratio and, indeed, in these two only can it be anything like constant. Travelers on the public highway and trespassers can only come into the enumeration when they become victims of casualty.

An examination of the detailed figures as shown in Table A is instructive, as bearing on this question. Consider, for example, the two classes, impossible to enumerate in total, referred to above. The percentage of the total number of casualties affecting travelers on the highway does not vary greatly in the two years 1891 and 1892, being 5 % in 1891 and 5.9 % in 1892. The actual number of casualties, however, was 190 in 1891 and 239 in 1892, a variation of over 25 %. Is there any reason for assuming that more travelers on the highway were *exposed* to injury in 1892 than in 1891? In 1892 the number of persons injured was 700 greater than in 1891. If the method of investigation now begun be maintained for a sufficient length of time, causes for such variations in the total and in the distribution will undoubtedly be discovered. Whatever might have been the cause, in the present instance, of the increased number of casualties, it looks very much as if the increased vigilance exercised over the safety of passengers had shunted a part of the hazard over to the employés and travelers on the public highway, although there is no marked increase of percentage in either case. The matter might be put in this way: Seven hundred more people will be injured this year than last; employés and travelers on the highway will get a little more than their share of the increase, because the corporation is going to take a lit-

TABLE B.—CLASSIFIED AS TO NATURE OF INJURY.

	NUMBER OF CASUALTIES.					RATIO OF DISTRIBUTION.				
	1891.	1892.	1893.	1894.	Mean.	1891.	1892.	1893.	1894.	Mean.
						%	%	%	%	%
Death	348	366	341	266	330	10.3	9.0	8.5	8.6	9.1
Loss of limb	90	90	84	71	84	2.7	2.2	2.1	2.3	2.3
Loss of finger or toe	101	121	105	79	102	3.0	3.0	2.7	2.6	2.8
Spinal injury	15	51	105	21	48	.4	1.3	2.7	.6	1.3
Fracture or dislocation	225	269	268	194	239	6.6	6.6	6.6	6.2	6.5
Sprains	369	426	362	411	392	10.9	10.5	9.1	13.5	11.0
Cuts and bruises	1522	1913	1893	1134	1615	44.9	47.2	47.4	36.6	44.0
Miscellaneous	720	811	836	918	821	21.2	20.2	20.9	29.6	23.0

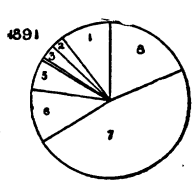

FIG. 6.

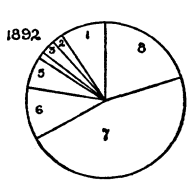

FIG. 7.

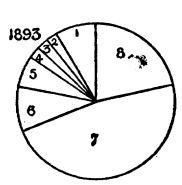

FIG. 8.

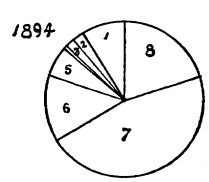

FIG. 9.

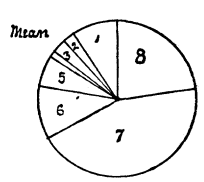

FIG. 10.

RESULTS.

1. Death.
2. Loss of limb.
3. Loss of finger or toe.
4. Spinal injury.
5. Fracture or dislocation.
6. Sprains.
7. Cuts and bruises.
8. Miscellaneous.

tle extra care of the passengers. The fourth class, the trespassers, seem, as usual, to have looked out for themselves and to have come out with a little less than their share of damage as shown by the previous year's experience.

The numbers for the year 1893 are interesting. This included a period of excessive passenger traffic, under conditions likely to considerably increase the total number of casualties. It was really, however, little different from and somewhat less than for the year 1892. Nor is the variation in ratio of distribution great, the several percentages agreeing well with the mean of the whole period.

But the most curious and interesting result brought out by the investigation is the constancy of ratio of distribution of *injuries among various classes*, such as death, loss of limb, loss of finger, fracture, etc. It is difficult to estimate the *a priori* probability of any one of these occurrences, and the facts here cited furnish a remarkable illustration of the operation of the principle or Law of the Long Run, as defined above. Indeed, in numbers relatively so small it is extremely surprising to find so many instances of persistency. General results as to character of injury and without reference to occupation are shown in diagrams 6, 7, 8, 9 and 10, and also numerically in Table B. We are here treating a total of about 15,000 casualties, an average of something less than 4,000 per annum. It would hardly have been expected in advance that during these four years there would be a nearly constant percentage of the total number of injured who would lose a limb, or that year after year almost exactly the same proportion would lose a finger or toe, or that the ratio of fractures and dislocations to the whole would be still more persistent, and this notwithstanding the fact that the total number of casualties would vary more than thirty per cent. Even when the analysis is carried much further there appears striking evidence of the same uniformity of distribution, although naturally there might be much less of it. In illustration of this I may cite the following: When the class of employés alone is considered and their injuries classified into the eight groups above referred to, in the two groups which contain the number of those who have suffered from loss of limb, or loss of finger or toe, injuries in which there is little chance of mistake in diagnosis, we find:

	Total Number.				Percentage of Total Number of Injuries.			
	1891	1892	1893	1894	1891	1892	1893	1894
Loss of limb..	34	35	34	26	1.4	1.1	1.1	1.2
Loss of finger or toe.....	89	110	94	74	3.2	3.4	3.4	3.3

Thus while the actual number of casualties varied considerably, the proportionate distribution remained extremely constant, particularly in the case of the loss of finger or toe. Although agreement in results where the numbers involved are not large must, itself, be regarded as fortuitous, it is interesting to note that in the class of trespassers, composed, it may be assumed, very largely of 'tramps,' of whom little regularity of any kind might be expected, and of whom about 450 are annually injured in one way or another, the ratio of loss of limb was in 1891, 10.8%; in 1892, 9.7%; in 1893, 9.4%, and in 1894, 9.8%. Of the same uncertain class, it is curious that in 1891, 8 suffered the loss of a finger or toe; in 1892, 9 suffered in the same way; in 1893, 8; but in 1894 this was reduced to 4.

Examples might be multiplied to almost any extent, but it is believed that enough has been shown to establish the existence, in this instance, of the principle under consideration. Indeed, so strong is the evi-

dence that we may feel quite justified in declaring that some error has crept into the classification of the injuries included in the three groups, sprains, cuts and bruises, and miscellaneous, as shown in the table for the year 1894. In short, it is more probable that error exists, either clerical, or arising from unusual professional carelessness in diagnosis, than that percentages of distribution, which have persisted so regularly during the three preceding years, should suddenly change to the extent shown in the table.

It is quite likely that the several accident insurance companies of the country have accumulated material relating to fortuitous events much more extensive than the above, which would yield equally interesting results if subjected to analysis.

There is one point to which it seems worth while to invite especial attention, namely, the confusion which often exists as to the inherent improbability of certain events. Such events are those which, for reasons entirely independent of the probability of their occurrence, have a particular interest. As an illustration, I may refer to the chance of the appearance of a particular hand at whist. Two or three years ago those interested in games with cards were greatly excited by the alleged occurrence of an event in the Boston & Albany railroad station in Boston. It was nothing less than that during the progress of a game of whist played by three railroad conductors and a mail agent, while waiting for the hour of departure of their trains, on taking up the cards after a deal each man found himself in possession of the whole thirteen cards of one suit. The *a priori* probability of such an event is all but infinitely small, and it was thought to be necessary to fortify the account published with affidavits of all the players and also of one or two gentlemen who happened to be watching the game. It probably occurred to few who read this account that the chances against any other particular distribution of the cards were just as great as against this, and that the result of every deal of the cards is just as remarkable as this and as little likely ever to occur again in the lifetime of the players. Indeed, any event of life, when considered in connection with contemporaneous and related events, in all their ramifications, will be found to have *a priori* chances so overwhelmingly against it that it seems impossible that it ever should happen. An 'accidental' death, for example, is an event generally unlikely, but in any specific case enough collateral circumstance and related fact can always be found to render the *a priori* probability of the combination nearly infinitely small. The chances of any man whom you may name meeting his death by falling from the third-story back window of the house belonging to his grandmother on his mother's side, and impaling himself on the point of a cotton umbrella accidentally left wide open in the garden below by the man servant of a gentleman named Witherspoon, temporarily stopping at the nearest inn, to whom he had loaned it on the day before at 2 P. M., in the lull of a thunderstorm which came from the north, are indefinitely small; yet I have been told that a man actually lost his life in just that way, and it is easy to see that the exact repetition of the simplest event in life, with all of its accompanying conditions and relations, would be just as incredible as this.

T. C. MENDENHALL.
WORCESTER POLYTECHNIC INSTITUTE.

HORTICULTURE AT CORNELL.

IN response to a request from the editor of SCIENCE, a brief outline of the purposes and methods of the work in horticulture at Cornell University is here given. This is the more willingly given because no full statement has been made of the capabilities